Data Assimilation

Alan O’Neill

University of Reading, UK

The Data Assimilation Problem

How can we combine noisy measurements of a system with output from an imperfect numerical model to get the best estimate of the (evolving) state of the system?

What are the benefits of data assimilation?

- Quality control
- Combination of data
- Errors in data and in model
- Filling in data poor regions
- Designing observing systems
- Maintaining consistency
- Estimating unobserved quantities
- Parameter estimation in models *****

Plan

- Motivation & basic ideas
- Univariate (scalar) data assimilation
- Multivariate (vector) data assimilation
  - 3d-Variational Method (& optimal interpolation)
  - Kalman Filter (+ extended KF)
  - Ensemble methods (+ particle filter)
  - 4d-Variational Method
- Applications of data assimilation in earth system science

In what follows, you don’t have to follow the mathematics to understand the basic ideas and to learn something useful!

Solving the Data Assimilation Problem

- The observations
- Their errors
- Predictions by a numerical model of the system
- The errors in these predictions

The key idea is to combine observations with predictions giving more weight to information with the least error. But errors may not be well known! Internal consistency checks on our state estimates are possible, but also need independent (unassimilated data).

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

Errors everywhere

All significant sources of uncertainty should be accounted for in data assimilation

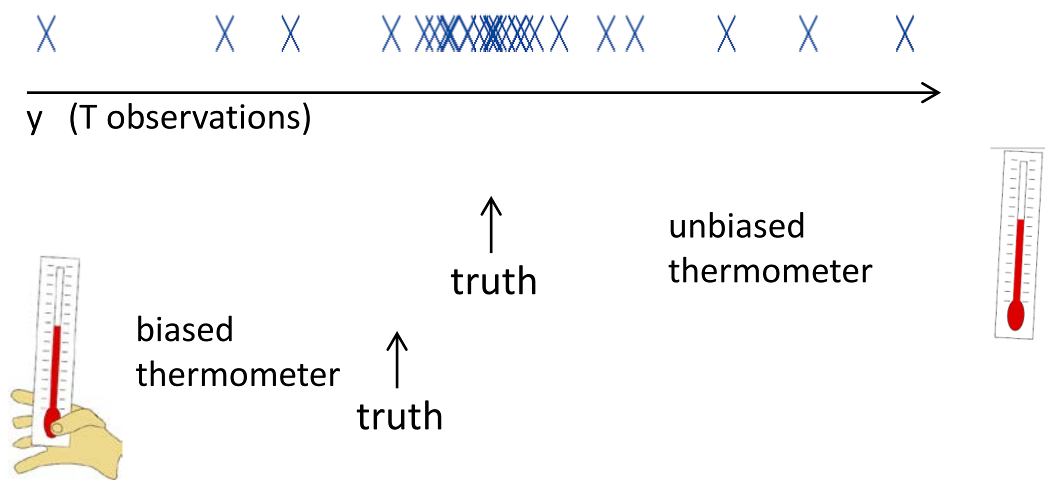
Random errors:

- background (a-priori) errors
- observation errors
- model errors
- representation errors

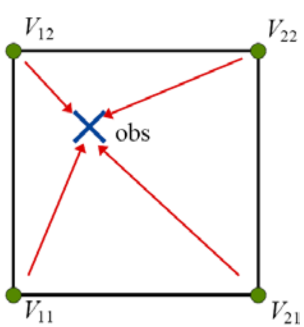
Systematic errors:

- biases in background
- biases in observations
- biases in model

Example 1 – repeated observations of air temperature



Example 2 – representation errors due to model grid



Best Linear Unbiased Estimate (BLUE)

Assume a numerical model predicts x to be x_b;
assume the observed value of x is y.

Make an estimate (the "LUE") of x as follows:

$\hat{x} = (1-\beta)x_b + \beta y$ linear, unbiased estimate (weights sum to 1)

$\hat{x}$ is the estimate of variable x;
x_b is the value of the variable x predicted by a model (the "prior", or "background");

y is the observed value of the variable x.

Think of these variables as RANDOM VARIABLES.
Assume statistical properties of their errors are known

BLUE

$$\hat{x} = x_b + \beta(y - x_b)$$

What should we choose for the value of β?

Choose β to minismise the variance of the error in the estimate:

$$< (\hat{x} - x_t)^2 >, \text{ where } x_t \text{ is the (unknown) true value of } x$$

This is what we mean by "best."

Note: we never need to know the true value of anything, just the statistics of the departures from the truth, i.e. the error statistics.

Important Points on Error Statistics in our treatment of Data Assimilation

- We assume that error statistics are Gaussian (either explicitly or implicitly). Thus we consider only the low order moments of the probability density functions (pdfs) of random variables: mean and variance (or standard deviation).
- We assume throughout that errors are unbiased (average = 0)
- These are important limitations:
  - Nonlinearity inevitably leads to “non Gaussianity”
  - Models almost always have some biases somewhere.

BLUE & observation operator

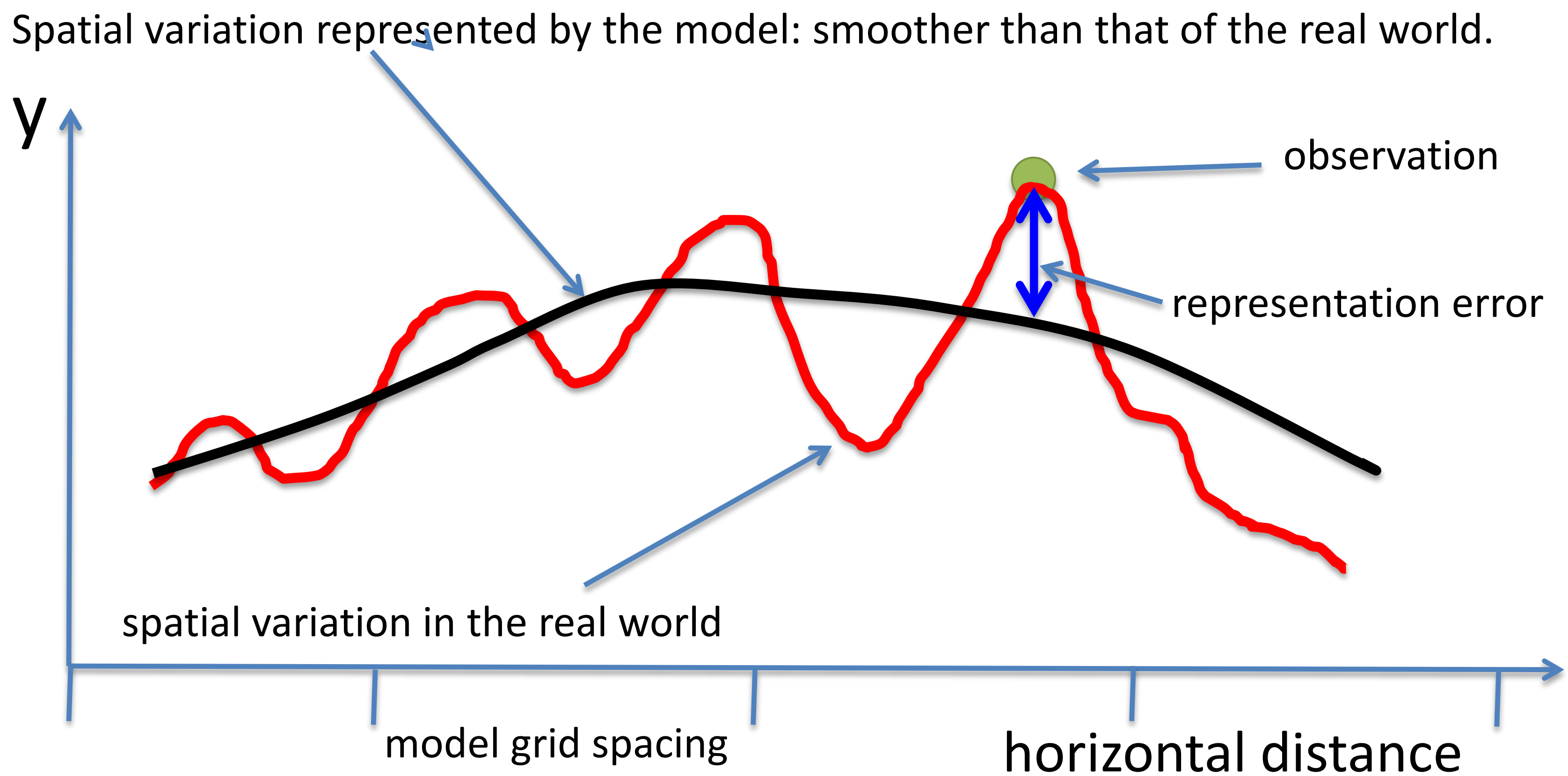
Assume observation is not a direct measurement of the variable used in our model.

For example, the model variable could be temperature, but the observation could be a radiance.

Let the relationship between the two be given by: $y = h(x)$, where h is called the observation operator.

In general, h is non linear, for example $y = \sigma x^4$, where x is temperature and y is radiance.

Representation Error



we are estimating the "smoothed state," so representation error associated with observation!!

BLUE (incl. obs. op.)

$$\hat{x} = x_b + \beta(y - h(x_b))$$

$x_b = x_t + \varepsilon_b$, where subsctipt t indicates (unkown) true value

$$y = y_t + \varepsilon_y$$

$$\hat{x} = x_t + \varepsilon_a$$

$$h(x_b) = h(x_t + \varepsilon_b) = h(x_t) + \frac{\partial h}{\partial x_t}\varepsilon_b = h(x_t) + H\varepsilon_b$$

where H is the "tangent linear" of the observation operator

Substitute into our estimate, and we find that

BLUE

$$\varepsilon_a = \beta\varepsilon_o + (1-\beta H)\varepsilon_b$$

where $\varepsilon_o = \varepsilon_y +$ error in h + representation error

$(y_t = h(x_t) +$ error in $h +$ representation error)

$$<(\hat{x}-x_t)^2> = <\varepsilon_a^2> = \beta^2 <\varepsilon_o^2> + (1-\beta H)^2 <\varepsilon_b^2>$$

where we have assumed errors are uncorrelated, $<\varepsilon_o\varepsilon_b> = 0$

$$\sigma_a^2 = \beta^2\sigma_o^2 + (1-\beta H)^2\sigma_b^2; \quad \mathrm{A} = \beta^2 R + (1-\beta H)^2 B$$

BLUE

Set $\frac{d}{d\beta}A = 0$ and solve for β

$$\beta_{\min} = W = \frac{HB}{R + H^2 B}$$

$$A = (1 - WH)B$$

Using the observation has given us a better estimate than we got from the model alone!

Let $H = 1$ for simplicity (y is a direct measurement of x)

$B \to \infty \quad \Rightarrow \quad \hat{x} \to y$ and $A \to R$

$R \to \infty \quad \Rightarrow \quad \hat{x} \to x_b$ and $A \to B$

BLUE with dynamics (toward the KF)

$x_b^{n+1} = M(\hat{x}^n)$ make a forecast to time n+1 starting from our estimate of the state at time n.

$x_b^{n+1} = M(x_t^n + \varepsilon_a^n) = M(x_t^n) + \mathrm{M}^n \varepsilon_a^n$ using Taylor expansion, where M^n is tangent linear model at time n

$x_b^{n+1} = x_t^{n+1} + \varepsilon_b^n + \mathrm{M}^n \varepsilon_a^n$

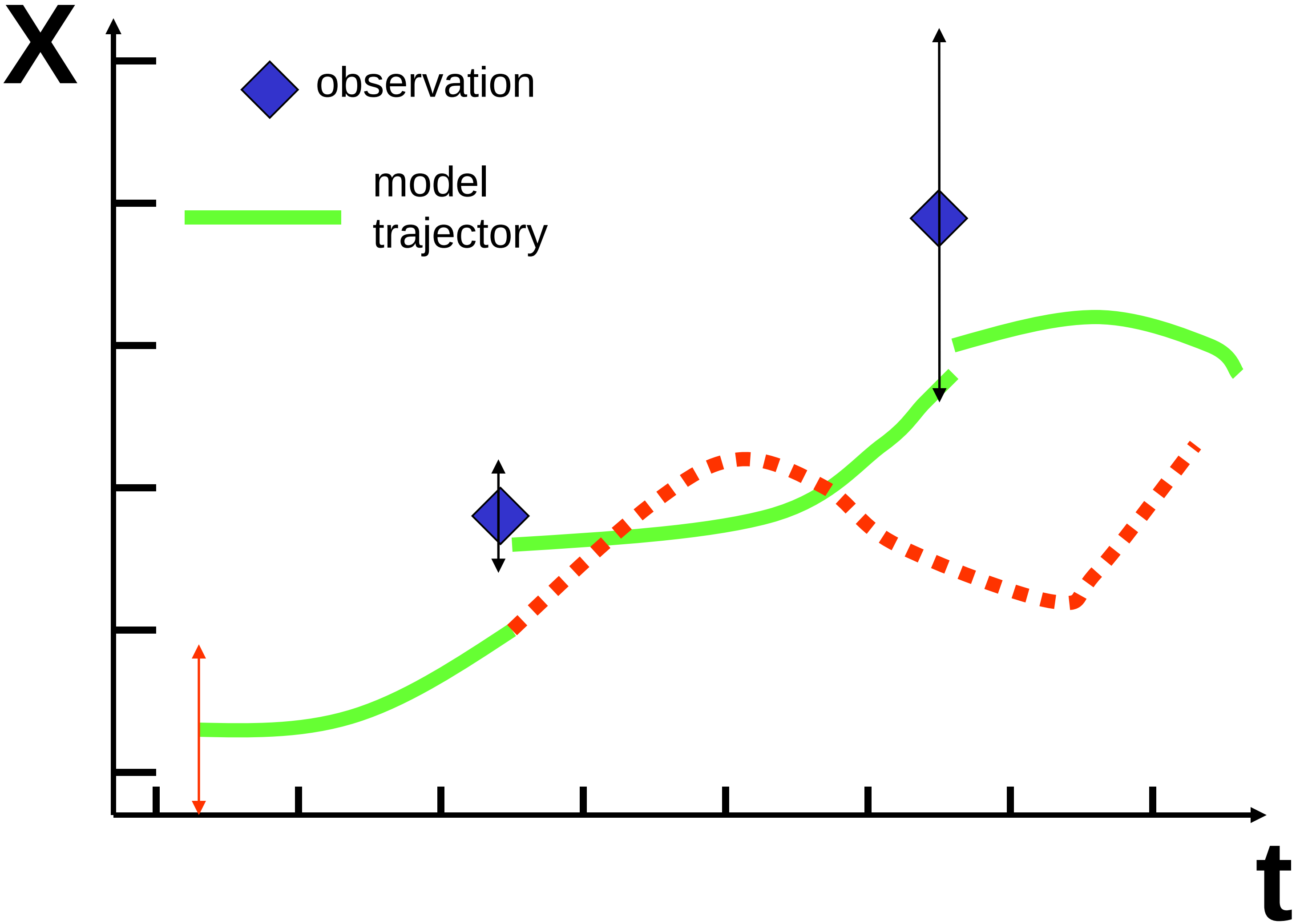
$\varepsilon_b^{n+1} = \varepsilon_b^n + \mathrm{M}^n \varepsilon_a^n$

$<(\varepsilon_b^{n+1})^2> = <(\varepsilon_b^n)^2> + (\mathrm{M}^n)^2 <(\varepsilon_a^n)^2>$ assuming error correlations=0

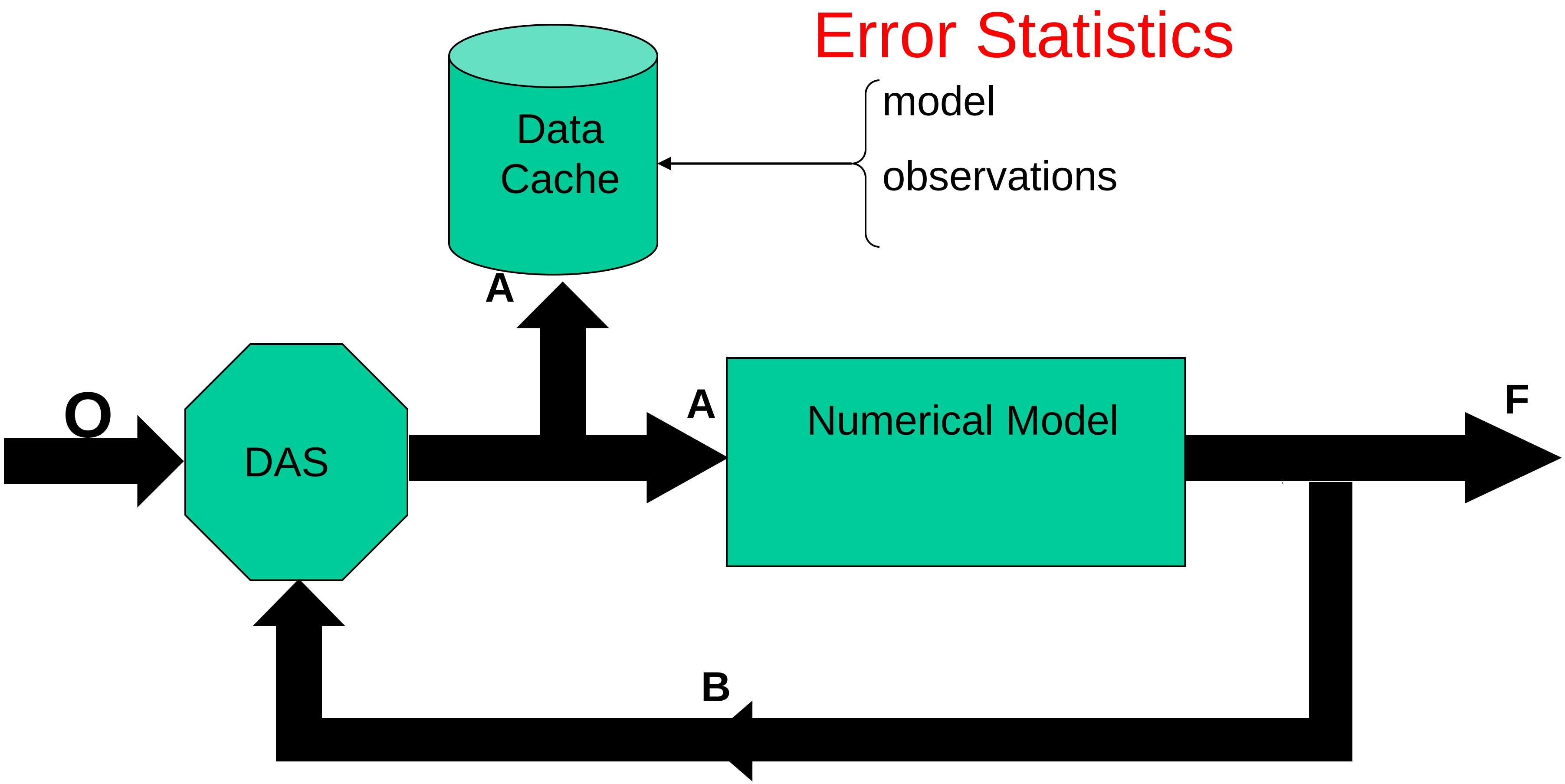
$B^{n+1} = B^n + (\mathrm{M}^n)^2 A^n$ eqn. governing change of forecast error

$R^{n+1} = R^n$ assume observation error does not change in time

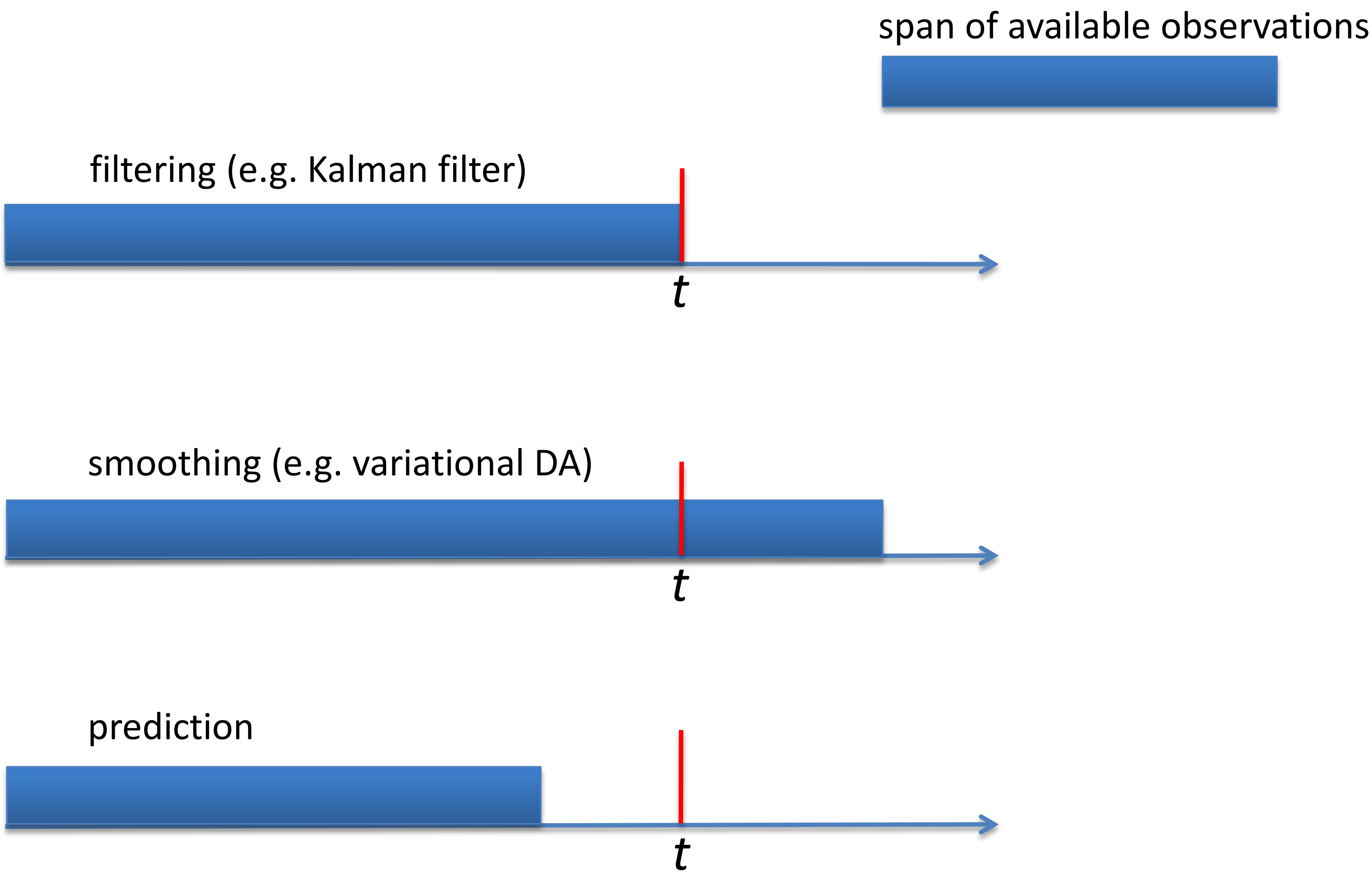
$A^{n+1} = (1 - W^{n+1} H) B^{n+1}$ eqn. governing change of estimate error


X
observation
model
trajectory
t

DATA ASSIMILATION SYSTEM



Three types of estimation problem (estimate desired at time t)



Rev. Thomas
Bayes
1702-1761

The Grand Unifying Principle in Data Assimilation

Bayes's Theorem (involving the concept of conditional probability)

Conditional Probability & Bayes' Theorem

$p(A,B) = p(A \mid B)\,p(B) = p(B \mid A)\,p(A),$

where A and B are two random events

Bayes' Theorem: $p(A \mid B) = \dfrac{p(B \mid A)\,p(A)}{p(B)}$

$$\therefore\ p(x \mid y) = \frac{p(y \mid x)\,p(x)}{p(y)},$$

where x is a state variable of the system we wish to estimate, and y is a measurement of that variable.

So if we have some prior information about $p(x)$, we can update that information with an observation y to get $p(x \mid y)$, the probability that the system variable has value x given that a measurement y of that variable has been made. We call it the posterior pdf.

Maximum Likelihood (minimizing a "cost function")

Assume we have an observation x_o of an unknown variable x.

Assume we have some prior information that the value of x is x_b.

Assume we know the error statistics of these quantities (the error variances).

$$p(x \mid x_o) \sim p(x_o \mid x) p(x) = \exp\left\{ -\frac{(x_o - x)^2}{\sigma_o^2} \right\} \exp\left\{ -\frac{(x - x_b)^2}{\sigma_b^2} \right\}$$

$$J(x) = -\ln p(x \mid x_o) \sim \left\{ \frac{(x_o - x)^2}{\sigma_o^2} \right\} + \left\{ \frac{(x - x_b)^2}{\sigma_b^2} \right\}, \text{ the COST FUNCTION}$$

Find x such that $J(x)$ is a minimum or such that $p(x \mid x_o)$ is a maximum,

i.e maximum probability or likelihood.

This x is our estimate $\hat{x}$.

$\hat{x} \sim \frac{x_o}{\sigma_o^2} + \frac{x_b}{\sigma_b^2}$. Easy to show from form of $p(x \mid x_0)$ that $\frac{1}{\sigma_a^2} = \frac{1}{\sigma_o^2} + \frac{1}{\sigma_b^2}$

The bigger the variance, the less weight is given to the information.

The precision of the estimate is better than those of the observation or background.

To get an equals sign in the above, divide by the sum of the weights. We get :

$$\hat{x} = x + W(x_o - x_b)$$

$$W = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2} \quad \text{and} \quad \sigma_a^2 = (1 - W)\sigma_b^2$$

So maximum likelihood method gives same results as minimum variance for Gaussian statistics.

BLUE or Optimal Interpolation (OI): multivariate case

System and observations must be represented by vectors not scalars.

Error variances become covariances, represented not as numbers but as matrices.

Error covariances represent the error correlations between different variables, e.g. temperature at different grid points, or temperature and wind at the same point.

Multivariate Case

state vector $$\mathbf{x}(t) = \begin{pmatrix} x_1 \\ x_2 \\ \bullet \\ \bullet \\ x_n \end{pmatrix}$$

observation vector $$\mathbf{y}(t) = \begin{pmatrix} y_1 \\ y_2 \\ \bullet \\ y_m \end{pmatrix}$$

Errors

The Error Covariance Matrix

$$\varepsilon = \begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix} \qquad \varepsilon^{\mathrm{T}} = \begin{pmatrix} e_1 & e_2 & . & . & . & e_n \end{pmatrix}$$

$$< e_i e_i > = \sigma_i^{\ 2}$$

$$\mathbf{P} = < \varepsilon \varepsilon^{\mathrm{T}} > = \begin{pmatrix} < e_1 e_1 > & < e_1 e_2 > & . & . & . & < e_1 e_n > \\ < e_2 e_1 > & < e_2 e_2 > & . & . & . & < e_2 e_n > \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ < e_n e_1 > & < e_n e_2 > & . & . & . & < e_n e_n > \end{pmatrix}$$

Background Errors

- They are the estimation errors of the background state (a model forecast):

$$\varepsilon_b = \mathbf{x}_b - \mathbf{x}$$

- Assume average (bias) $<\varepsilon_b>= 0$
- covariance

$$\mathbf{B} =< \varepsilon_b \varepsilon_\mathrm{b}^\mathrm{T} >$$

The Observation Operator

The observations (observation vector) are in general not direct measurements of the state variables (state vector), e.g. in remote sensing from space.
In data assimilation, we need to compare the observation vector with the state vector. The observation operator allows this.
It is a mapping from state space to observation space.

$$\mathbf{y}^{\mathrm{mod}} = h(\mathbf{x})$$

$$R_i = \int B_i(T(p)) \frac{d\tau}{dp}$$

Data assimilation algorithms often use the matrix evaluated generally at a state forecast by the model (background state or first-guess state)

$$\mathbf{H} = \left. \frac{\partial h}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_B}$$

Observation Errors

They contain:

- errors in the observation process (instrumental error),
- errors in the design of h, and
- "representation errors"

$\varepsilon_{\mathrm{o}} = \mathbf{y} - h(\mathbf{x})$, where $\mathbf{x}$ is the true state

$\mathbf{R} = < \varepsilon_{\mathrm{o}} \varepsilon_{\mathrm{o}}^{\mathrm{T}} >$ assuming no bias

Optimal Interpolation (the BLUE)

- BLUE = Best linear unbiased estimate
- Algorithm derived as a special case of 3D-var.

Algorithms require statistics of model/observation comparison

$$\mathbf{y} - h(\mathbf{x}_b)$$

$$\varepsilon = \mathbf{y} - h(\mathbf{x}_b) = \mathbf{y} - h(\mathbf{x} + \mathbf{x}_b - \mathbf{x})$$

$$= \mathbf{y} - h(\mathbf{x}) + \mathbf{H}(\mathbf{x}_b - \mathbf{x})$$

$$= \varepsilon_\mathrm{o} + \mathbf{H}\varepsilon_b$$

$$< \varepsilon\varepsilon^\mathrm{T} > = < \varepsilon_\mathrm{o}\varepsilon_\mathrm{o}{}^\mathrm{T} > + < \mathbf{H}\varepsilon_b\varepsilon_b{}^\mathrm{T}\mathbf{H}^\mathrm{T} >$$

$$= \mathbf{R} + \mathbf{HBH}^\mathrm{T}$$

BLUE Estimator (recursive)

- The BLUE estimator or "analysis" is given by:

$$\mathbf{x}_{\mathrm{a}} = \mathbf{x}_{\mathrm{b}} + \mathbf{K}(\mathbf{y} - h(\mathbf{x}_{\mathrm{b}}))$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{B}\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}$$

- The $\mathbf{B}$ matrix plays a key role in determining the structure of the analysed fields.
- Matrix inverses expensive to compute so reduce dimension by "local analysis"
- We can derive an explicit expression for the analysis error covariance matrix:

$$\mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}$$

Assumptions Used in *BLUE*

- Linearized observation operator:

$$\mathbf{h}(\mathbf{x}) - \mathbf{h}(\mathbf{x}_\mathrm{b}) = \mathbf{H}(\mathbf{x} - \mathbf{x}_\mathrm{b})$$

- Errors are unbiased:

$$< \mathbf{x}_\mathrm{b} - \mathbf{x} >=< \mathbf{y} - \mathbf{h}(\mathbf{x}) >= 0$$

- Errors are uncorrelated:

$$< (\mathbf{x}_\mathrm{b} - \mathbf{x})(\mathbf{y} - \mathbf{h}(\mathbf{x}))^\mathrm{T} >= 0$$

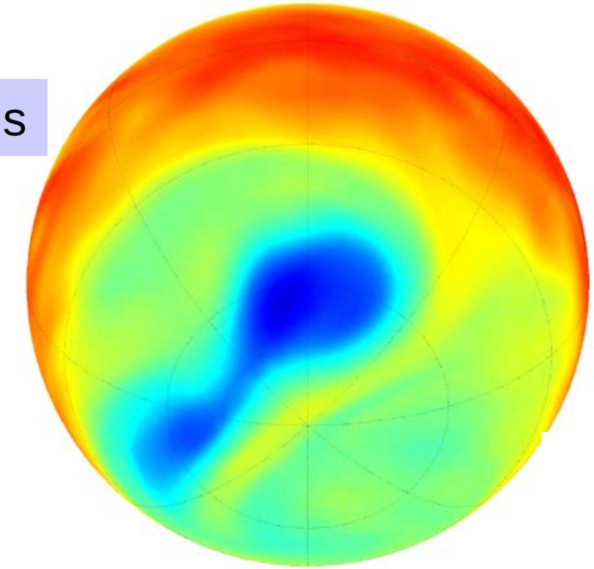
Innovations and Residuals

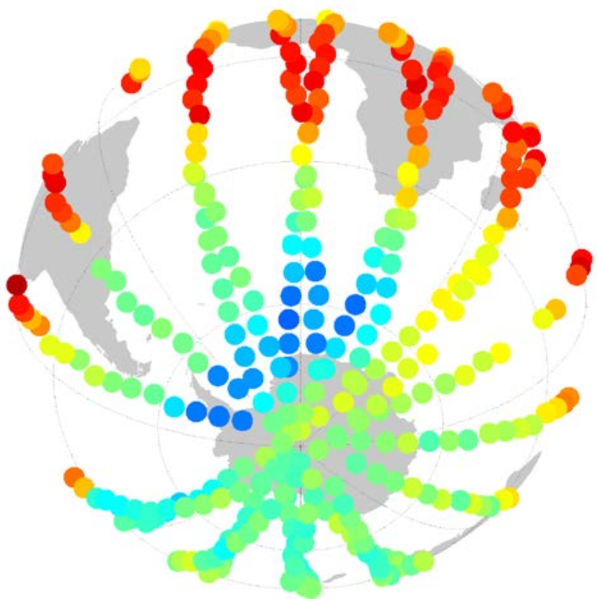
- Key to data assimilation is the use of differences between observations and the state vector of the system
- We call $\mathbf{y} - \mathbf{h}(\mathbf{x}_b)$ the innovation
- We call $\mathbf{y} - \mathbf{h}(\mathbf{x}_a)$ the analysis residual

Give important information

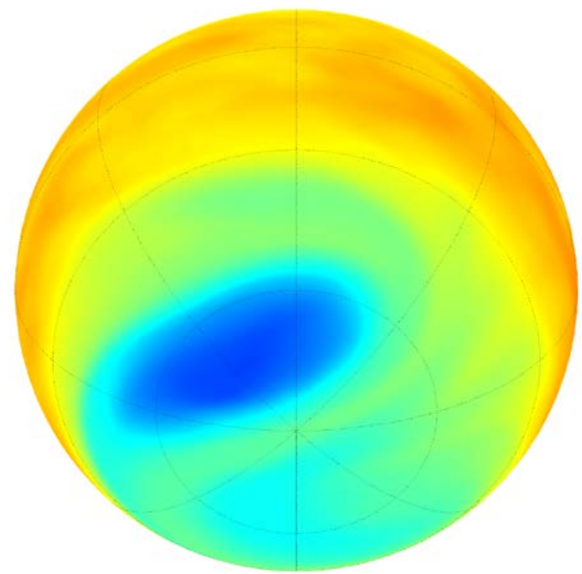
Ozone at 10hPa, 12Z 23rd Sept 2002

Analysis





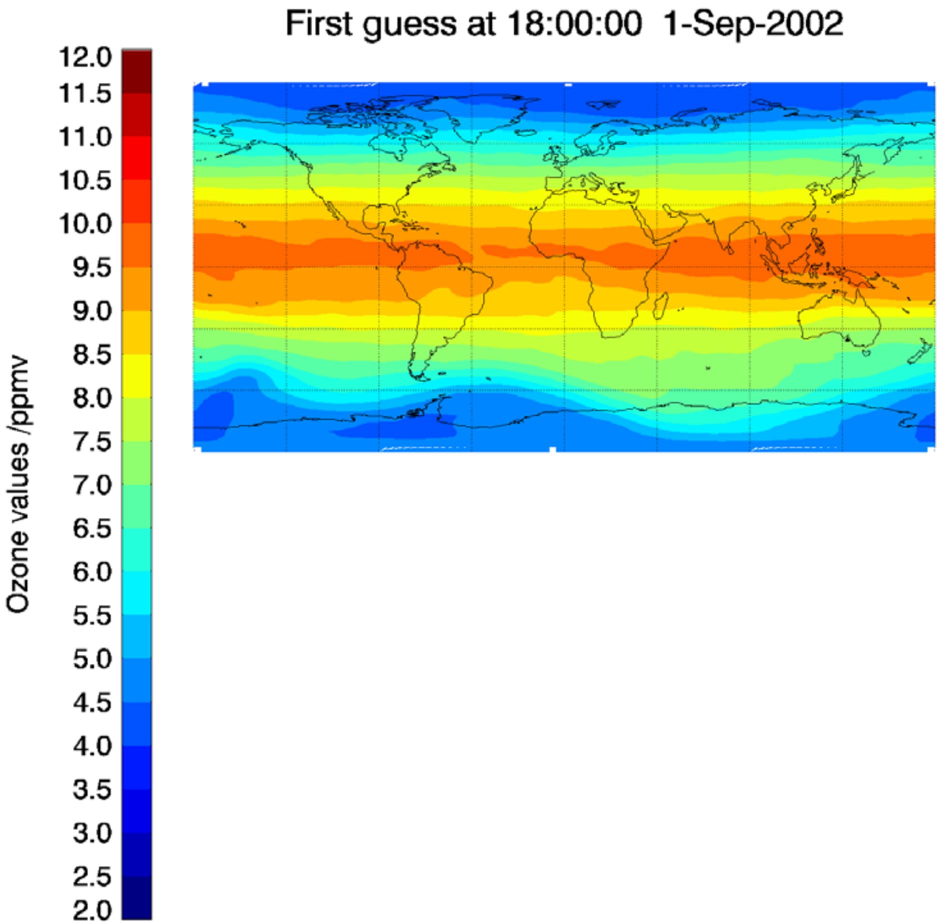
MIPAS observations



6 day model forecast

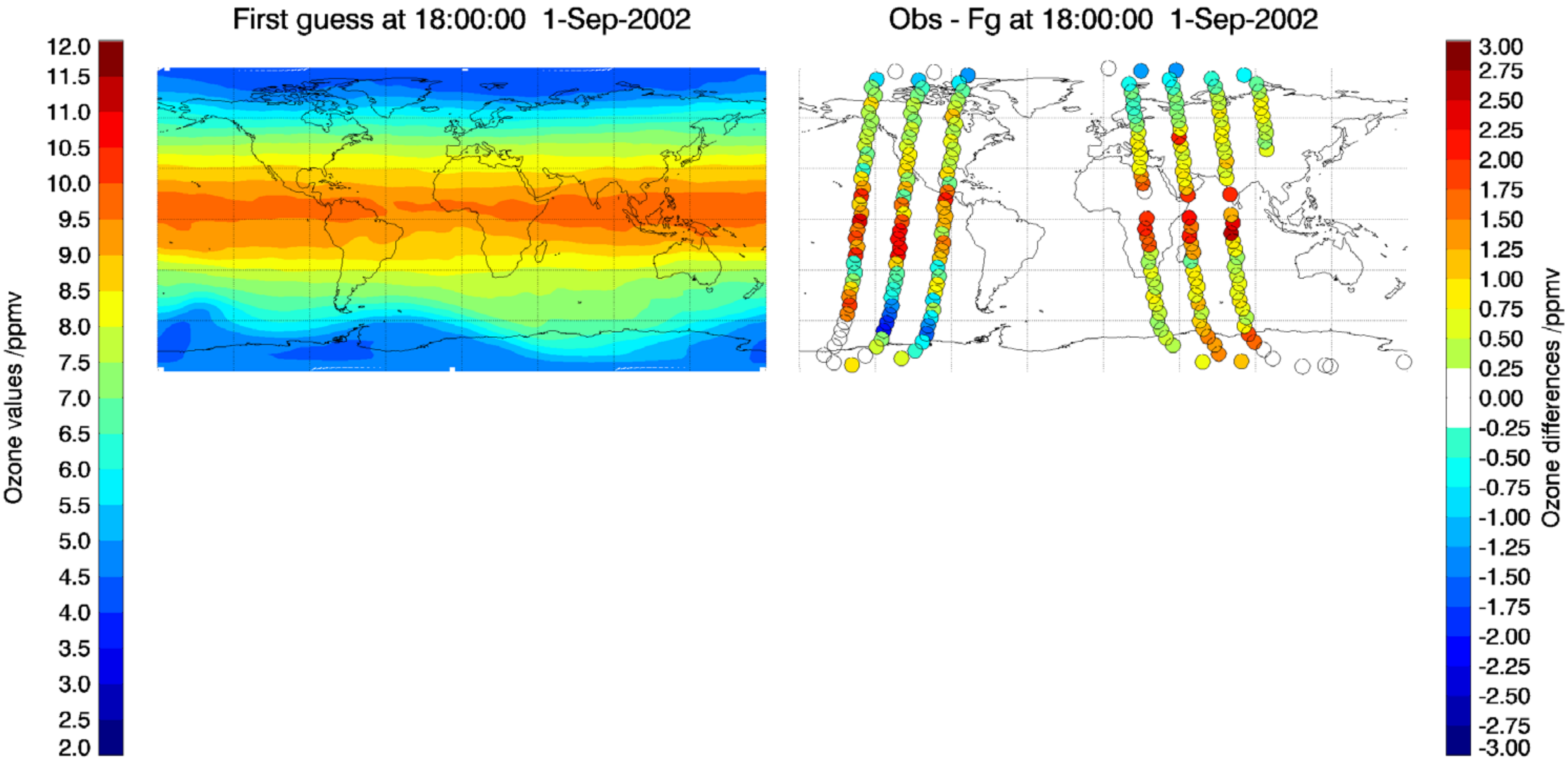
3D variational data assimilation - ozone at 10hPa

$\mathbf{x}_b$



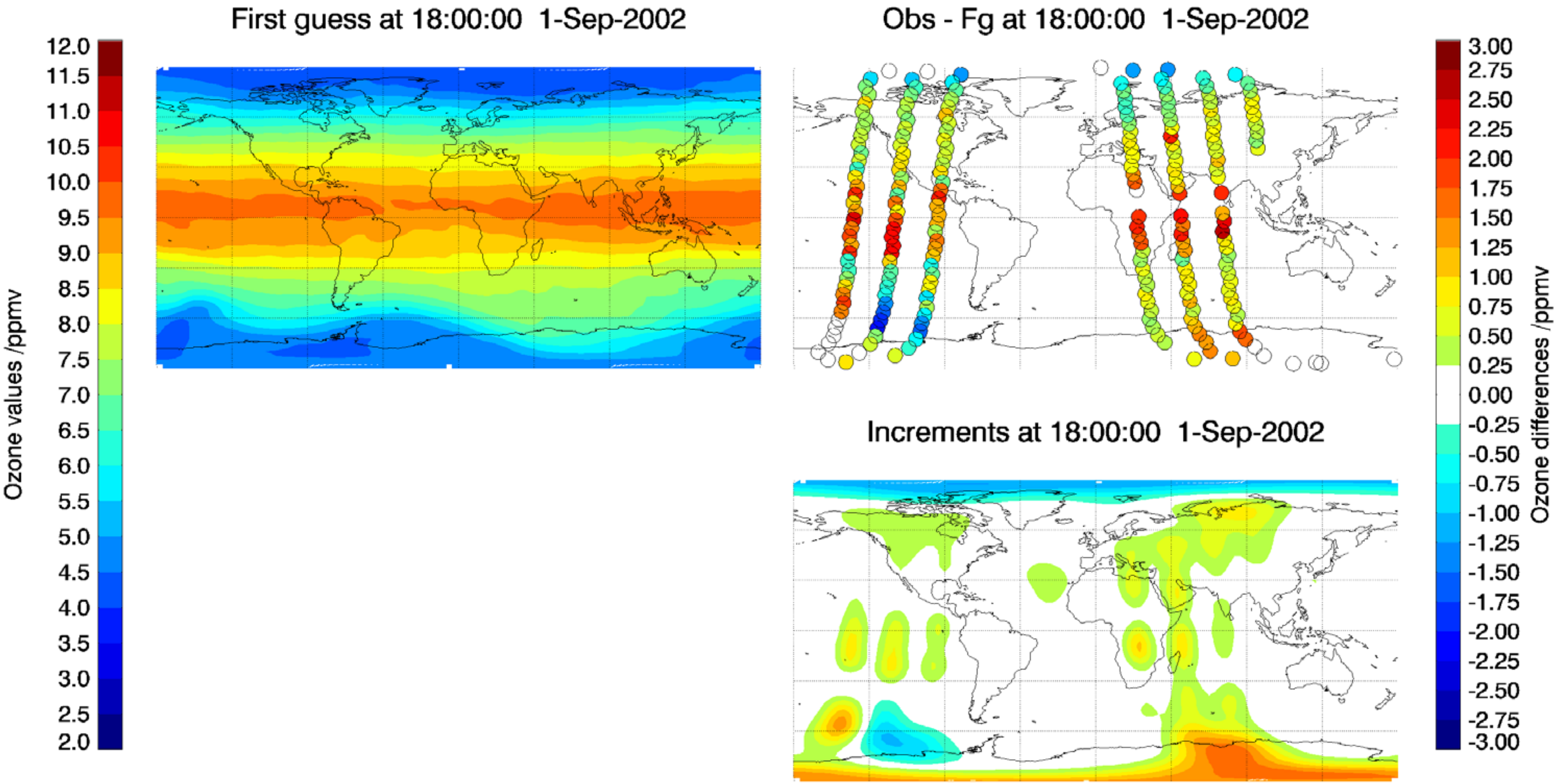
3D variational data assimilation - ozone at 10hPa

$\mathbf{x}_b$ $\qquad\qquad$ $\mathbf{y} - h(\mathbf{x}_b)$



3D variational data assimilation - ozone at 10hPa

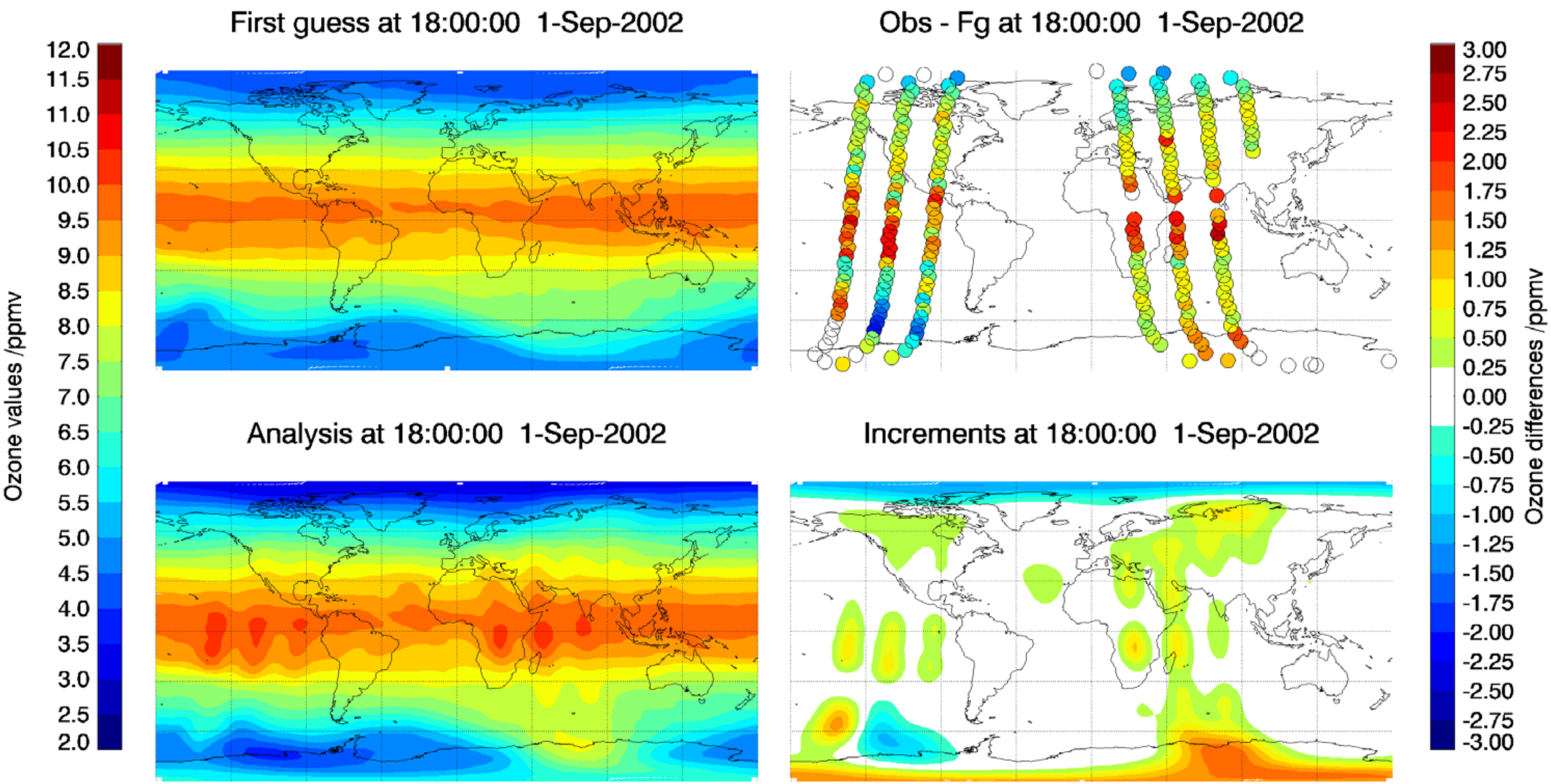
$\mathbf{x}_b$ $\qquad$ $\mathbf{y} - h(\mathbf{x}_b)$



$\mathbf{K}(\mathbf{y} - h(\mathbf{x}_b))$

The data assimilation cycle: ozone at 10hPa

$\mathbf{x}_b$ $\qquad\qquad$ $\mathbf{y} - h(\mathbf{x}_b)$



$\mathbf{x}_b + \mathbf{K}(\mathbf{y} - h(\mathbf{x}_b))$ $\qquad\qquad$ $\mathbf{K}(\mathbf{y} - h(\mathbf{x}_b))$

3D Variational Data Assimilation (3D-Var)

The Bayesian View of DA

$$p(\mathbf{x}\,|\,\mathbf{y}) = \frac{p(\mathbf{y}\,|\,\mathbf{x})\,p(\mathbf{x})}{p(\mathbf{y})} \propto p(\mathbf{y}\,|\,\mathbf{x})\,p(\mathbf{x})$$

$$p(\mathbf{x}\,|\,\mathbf{y}) \propto \exp\{-\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)\}\ \exp\{-\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{x}))\}$$

$$p(\mathbf{x}\,|\,\mathbf{y}) \propto \exp-\{\frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)+\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{x}))\}$$

Maximum likelihood: find $\mathbf{x}$ such that $p(\mathbf{x}\,|\,\mathbf{y})$ is a maximum.

Equivalently, define cost or penalty function (a scalar) by:
$J(\mathbf{x}) = -\ln p(\mathbf{x}\,|\,\mathbf{y})$.

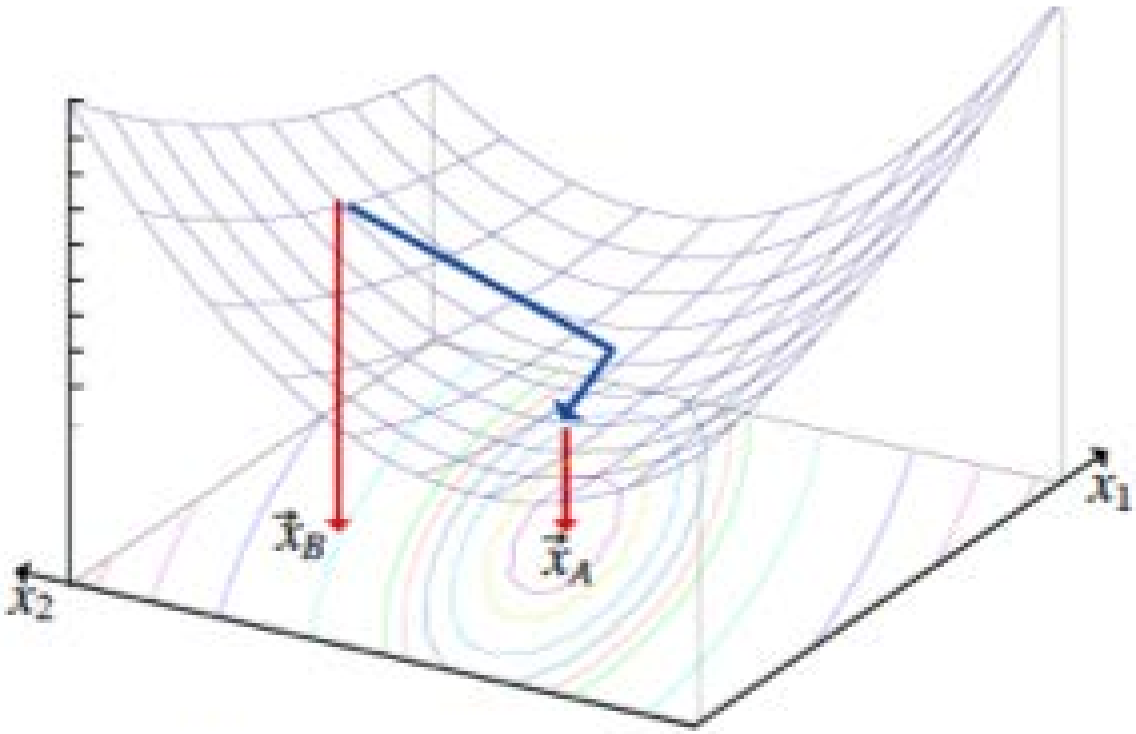
$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b)+\frac{1}{2}(\mathbf{y}-\mathbf{h}(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{h}(\mathbf{x}))$$

Find $\mathbf{x}$ to minimise the cost function

Minimizing the Cost Function

The problem involves a (badly constrained) optimization problem in 10^7 dimensions!

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \frac{1}{2}(\mathbf{y} - \mathbf{h}(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}))$$



Algebraic Minimization of the Cost Function

Assume it is reasonable to linearise the forward model $\mathbf{h}$.

$$\mathbf{h}(\mathbf{x}) \approx \mathbf{h}(\mathbf{x}_b) + \mathbf{H}(\mathbf{x} - \mathbf{x}_b)$$

Then we can show that $\partial J / \partial \mathbf{x} = \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}))$

Setting $\partial J / \partial \mathbf{x} = 0$ gives

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{B}\mathbf{H}^T(\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\mathbf{y} - \mathbf{h}(\mathbf{x}_b))$$

This is the same as the Optimal Interpolation or BLUE formula!

Algebraic minimization of the cost function

Under simplified conditions the cost function can be minimized algebraically.

Assume that the linearization of the forward model is reasonable

$$\vec{h}[\vec{x}] \approx \vec{h}[\vec{x}_B] + \mathbf{H}(\vec{x} - \vec{x}_B)$$

$$J[\vec{x}] = \frac{1}{2}(\vec{x} - \vec{x}_B)^T \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \frac{1}{2}(\mathbf{H}(\vec{x} - \vec{x}_B) - (y - \vec{h}[\vec{x}_B]\vec{)})^T \mathbf{R}^{-1}(\mathbf{H}(\vec{x} - \vec{x}_B) - (y - \vec{h}[\vec{x}_B]\vec{)})$$

1. Calculate the gradient vector

$$\nabla_{\vec{x}} J = \begin{pmatrix} \partial J / \partial x_1 \\ \partial J / \partial x_2 \\ \\ \partial J / \partial x_N \end{pmatrix} = \mathbf{B}^{-1}(\vec{x} - \vec{x}_B) + \mathbf{H}^T \mathbf{R}^{-1}(\vec{h}[\vec{x}] - \vec{y})$$

2. The special $\vec{x}$ that has zero gradient minimizes J (this cost function is quadratic and convex)

$$\nabla_{\vec{x}} J \big|_{x_A} = 0$$

$$\vec{x}_A = \vec{x}_B + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}(\vec{y} - \vec{h}[\vec{x}_B])$$

$$= \vec{x}_B + \mathbf{B}\mathbf{H}^T (\mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T)^{-1}(\vec{y} - \vec{h}[\vec{x}_B])$$

This is the OI formula with the BLUE!

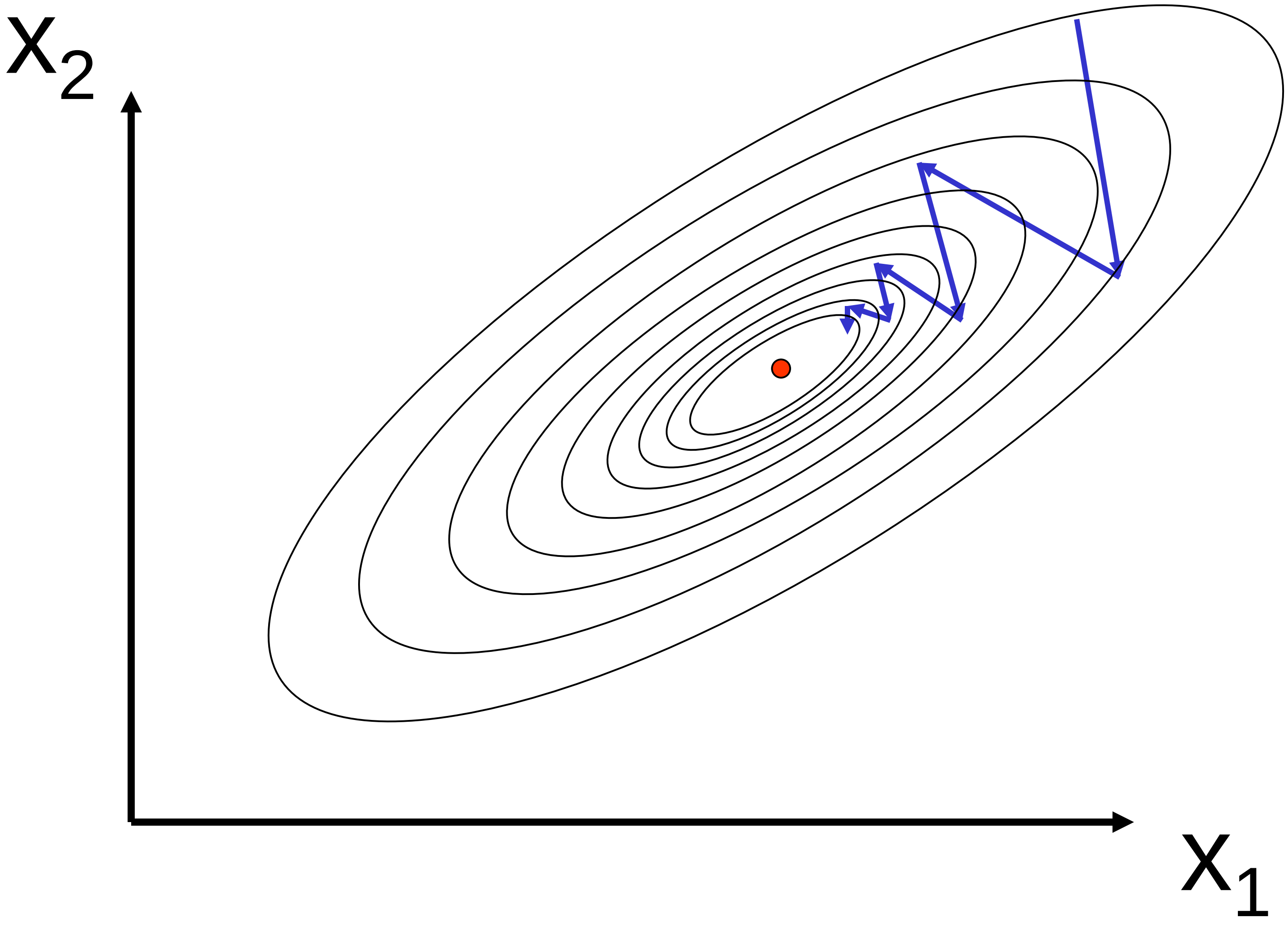
Remarks on 3d-VAR

- Can add constraints to the cost function, e.g. to help maintain “balance”
- Can work with non-linear observation operator *H.*
- Can assimilate radiances directly (simpler observational errors).
- Can perform global analysis instead of OI approach of radius of influence.

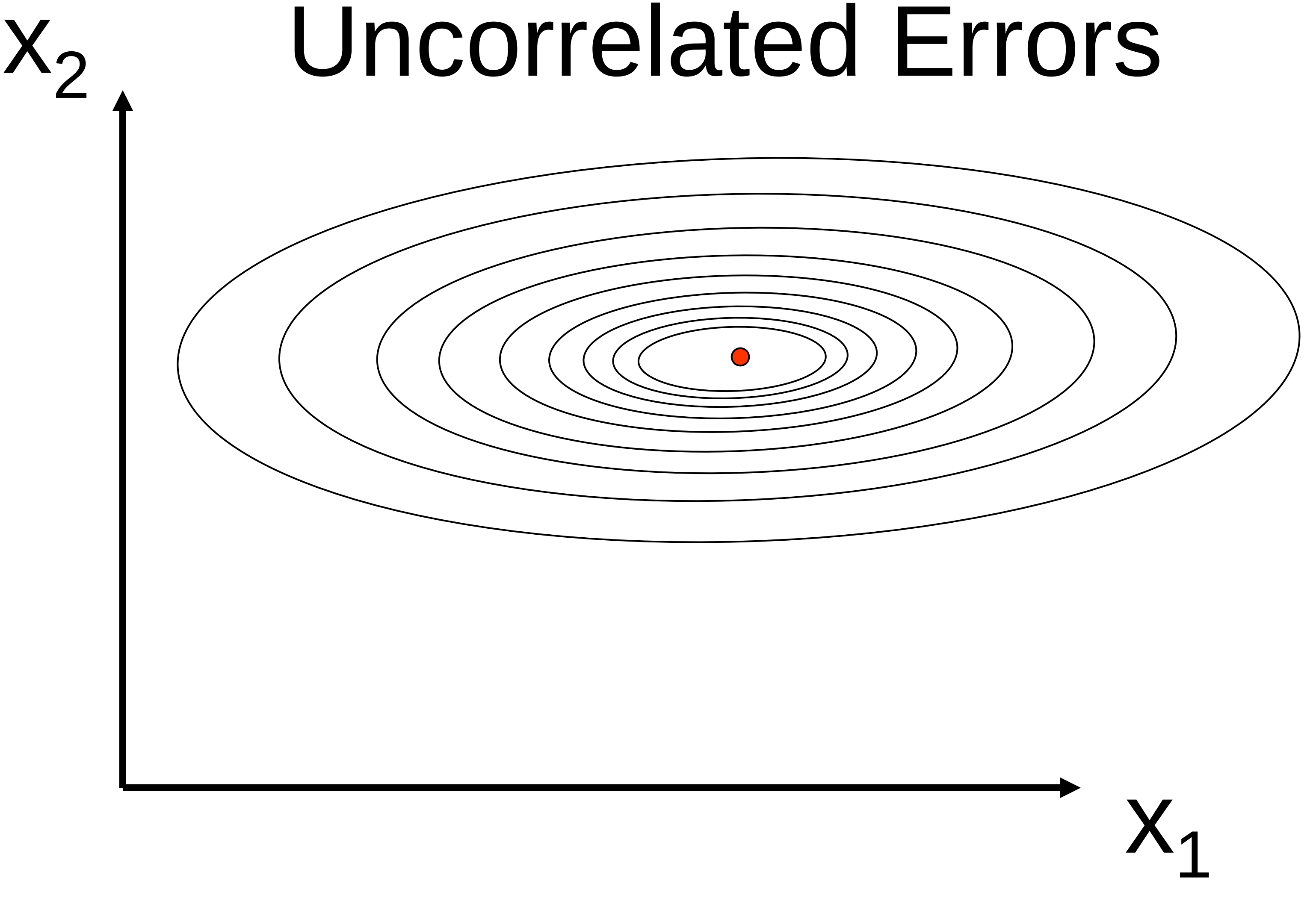
Choice of State Variables and Preconditioning

- Free to choose which variables to use to define state vector, $x(t)$
- We'd like to make B diagonal
  - may not know *covariances* very well
  - want to make the minimization of J more efficient by "preconditioning": transforming variables to make surfaces of constant J nearly spherical in state space

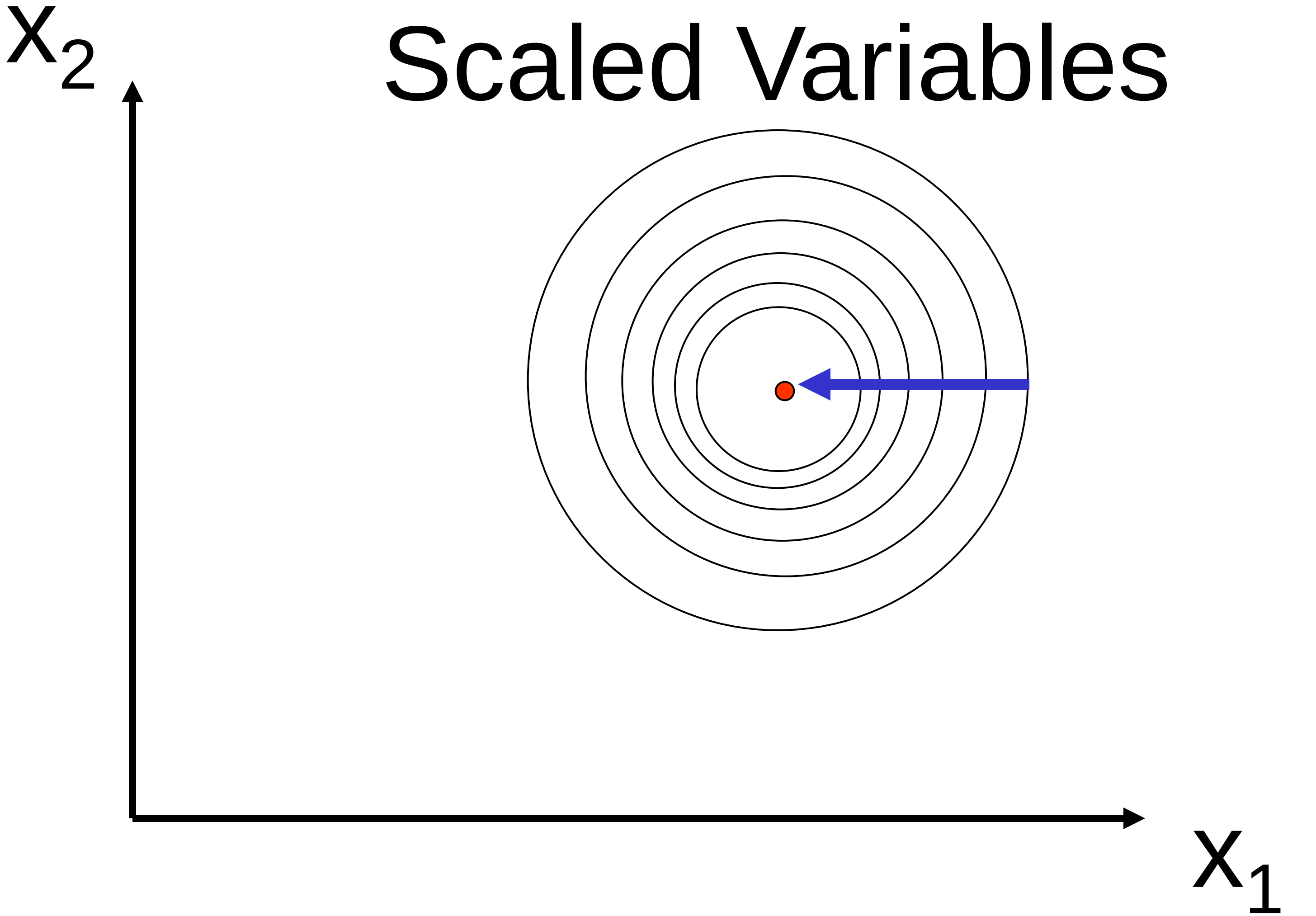
Cost Function for Correlated Errors



Cost Function for Uncorrelated Errors



Cost Function for Uncorrelated Errors Scaled Variables



END